



ТИНЬКОФФ

Визуализация данных

Зачем вам эта
лекция?

Пример отчета из письма

Согласие на оформление нового продукта в прошлом месяце



■ Да ■ Нет

Хорошо это или плохо?

Зачем вам эта лекция?

Согласие на оформление нового продукта в прошлом месяце

i Сколько клиентов было в тесте?
Вдруг 2?

i Как выглядит конверсия в
динамике?

i Почему отказываются?



i 50% - это итог или незрело?

i Верно ли логируются данные?

О чем эта лекция?



Визуализация – что это такое и зачем она нужна?



Как визуализировать? + разбор плохих примеров



Инструменты визуализации в Тинькофф



Немного практики на Python

Зачем нужна визуализация?

Центр сбора	Доход	Семейно е положен ие	Образова ние	Кол-во членов семьи	Количес тво детей	пол	Первый платеж	Должност ь	Номер договора	Заемщик _ID	Дата выдачи договора	Адрес рабочий	Адрес регистра ции	Адрес залога	Адрес фактичес кий	Номер телефона основной	Номер телефона домашни й	Номер телефона рабочий	Номер телефона дополнит ельный	Валюта договора
Москва	30000	Женат	Среднее с	2	0	М		Электром	624/0000-1	1-DO25K-7	17.11.2015	1	1	0	0	1	0	1	0	RUR
Москва	270000	Женат	Высшее	4	2	М	5187,92	коммерче	625/2200-1	1-63D5G-1	03.03.2014	1	1	0	0	1	0	0	0	RUR
Москва	270000	Женат	Высшее	4	2	М		коммерче	633/2703-1	1-63D5G-1	09.02.2012	1	1	0	0	1	0	0	0	RUR
Москва	40000	Женат	Среднее	3	1	Ж	0		625/0000-1	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	40000	Женат	Среднее	3	1	Ж			633/0203-1	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	30000	Не женат	Высшее	1	0	Ж	4062,7	Юрист	625/2000-1	1-PG-42	14.11.2005	1	1	0	0	1	1	0	0	RUR
Москва	87290	Не женат	Высшее	1	0	М		консульта	624/0000-1	1-9JVII-14	22.12.2009	1	1	0	1	0	1	1	1	RUR
Москва	400000	Женат	Дополнит	4	0	М	15281,08	первый за	625/3703-1	1-64VSL-6	10.11.2014	0	1	0	0	1	1	0	0	RUR
Москва	120169	Женат	Высшее	4	2	М		командир	624/0000-1	1-9SV6I-3I	25.12.2007	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М	1229,41	командир	625/1203-1	1-9SV6I-3I	28.12.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М		командир	633/0000-1	1-9SV6I-3I	18.04.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М	3889,8	командир	625/1203-1	1-9SV6I-3I	29.07.2014	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М	3486,11	командир	625/3500-1	1-9SV6I-3I	22.12.2011	0	1	0	0	0	0	1	1	RUR
Москва	30450	Не женат	Высшее	1	0	Ж	3675,78	оператор	625/2200-1	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	30450	Не женат	Высшее	1	0	Ж	1018,35	оператор	625/2200-1	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	50000	Разведен				М	0	Руководит	625/0000-1	1-DTQH-3	10.06.2015	0	1	0	0	1	1	1	1	RUR
Москва	48074	Не женат	Высшее	1	0	Ж		Менедже	633/2810-1	1-BKUP4-6	17.07.2012	1	1	0	0	1	1	0	0	RUR
Москва	48074	Не женат	Высшее	1	0	Ж		Менедже	624/0000-1	1-BKUP4-6	03.12.2012	1	1	0	0	1	1	0	0	RUR
Москва	87000	Не женат	Среднее с	1	0	М		МЕНЕДЖЕ	633/2610-1	1-BYW4Y-6	12.05.2014	1	1	0	1	1	0	0	0	RUR
Калуга (Вс	24641,23	Вдовец(в	Среднее с	1	0	Ж	0	Санитарк	625/0000-1	1-7FGOL-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Калуга (Вс	24641,23	Вдовец(в	Среднее с	1	0	Ж		Санитарк	633/1103-1	1-7FGOL-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Тамбов (Вс	850000	Женат	Высшее	2	0	М	0	Заместите	625/0000-1	1-7TWZ1-1	27.03.2012	0	0	0	0	1	1	1	1	RUR

Зачем нужна визуализация?

Центр сбора	Доход	Семейно е положен ие	Образова ние	Кол-во членов семьи	Количес тво детей	пол	Первый платеж	Должност ь	Номер договора	Заемщик _ID	Дата выдачи договора	Адрес рабочий	Адрес регистра ции	Адрес залога	Адрес фактичес кий	Номер телефона основной	Номер телефона домашни й	Номер телефона рабочий	Номер телефона дополнит ельный	Валюта договора
Москва	30000	Женат	Среднее с	2	0	М		Электром	624/0000-1	1-DO25K-7	17.11.2015	1	1	0	0	1	0	1	0	RUR
Москва	270000	Женат	Высшее	4	2	М	5187,92	коммерче	625/2200-1	1-63D5G-1	03.03.2014	1	1	0	0	1	0	0	0	RUR
Москва	270000	Женат	Высшее	4	2	М		коммерче	633/2703-1	1-63D5G-1	09.02.2012	1	1	0	0	1	0	0	0	RUR
Москва	40000	Женат	Среднее	3	1	Ж	0		625/0000-1	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	40000	Женат	Среднее	3	1	Ж			633/0203-1	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	30000	Не женат	Высшее	1	0	Ж	4062,7	Юрист	625/2000-1	1-PG-42	14.11.2005	1	1	0	0	1	1	0	0	RUR
Москва	87290	Не женат	Высшее	1	0	М		консульта	624/0000-1	1-9JVII-14	22.12.2009	1	1	0	1	0	1	1	1	RUR
Москва	400000	Женат	Дополнит	4	0	М	15281,08	первый за	625/3703-1	1-64VSL-6	10.11.2014	0	1	0	0	1	1	0	0	RUR
Москва	120169	Женат	Высшее	4	2	М		командир	624/0000-1	1-9SV6I-3I	25.12.2007	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М	1229,41	командир	625/1203-1	1-9SV6I-3I	28.12.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М		командир	633/0000-1	1-9SV6I-3I	18.04.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М	3889,8	командир	625/1203-1	1-9SV6I-3I	29.07.2014	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2	М	3486,11	командир	625/3500-1	1-9SV6I-3I	22.12.2011	0	1	0	0	0	0	1	1	RUR
Москва	30450	Не женат	Высшее	1	0	Ж	3675,78	оператор	625/2200-1	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	30450	Не женат	Высшее	1	0	Ж	1018,35	оператор	625/2200-1	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	50000	Разведен				М	0	Руководит	625/0000-1	1-DTTQH-3	10.06.2015	0	1	0	0	1	1	1	1	RUR
Москва	48074	Не женат	Высшее	1	0	Ж		Менедже	633/2810-1	1-BKUP4-6	17.07.2012	1	1	0	0	1	1	0	0	RUR
Москва	48074	Не женат	Высшее	1	0	Ж		Менедже	624/0000-1	1-BKUP4-6	03.12.2012	1	1	0	0	1	1	0	0	RUR
Москва	87000	Не женат	Среднее с	1	0	М		МЕНЕДЖЕ	633/2610-1	1-BYW4Y-6	12.05.2014	1	1	0	1	1	0	0	0	RUR
Калуга (Вс	24641,23	Вдовец(в	Среднее с	1	0	Ж	0	Санитарк	625/0000-1	1-7FGOL-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Калуга (Вс	24641,23	Вдовец(в	Среднее с	1	0	Ж		Санитарк	633/1103-1	1-7FGOL-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Тамбов (Вс	850000	Женат	Высшее	2	0	М	0	Заместите	625/0000-1	1-7TWZ1-1	27.03.2012	0	0	0	0	1	1	1	1	RUR

И что?

Зачем нужна визуализация?



Анализ данных

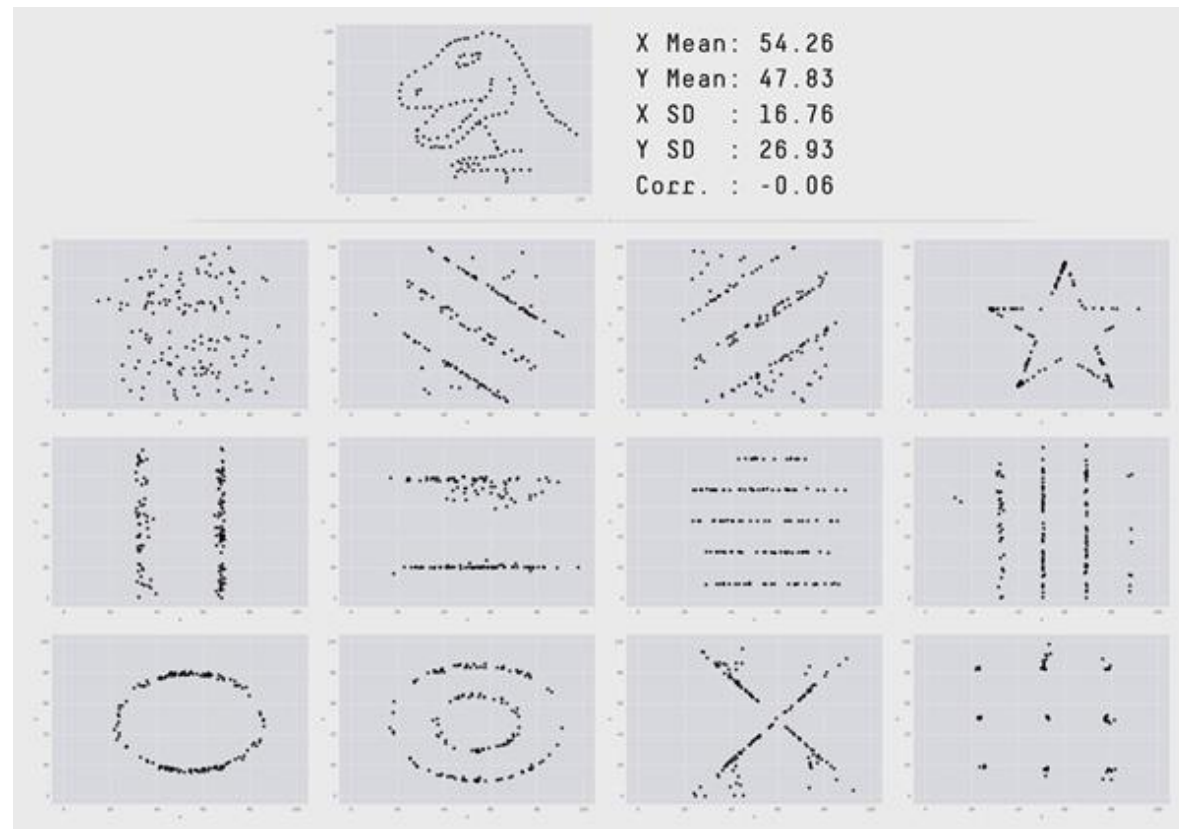


Регулярная отчетность, мониторинг процессов



Презентации

Дюжина Датазавров



Виды сравнения метрик

01

Покомпонентное

50,5% всего земного населения - это мужчины.

02

Позиционное

Средний размер арбуза больше среднего размера любой другой ягоды.

03

Временное

С 1800-ых годов температура Земли выросла на 1,1°C.

04

Частотное

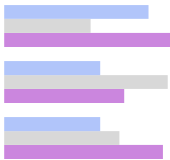
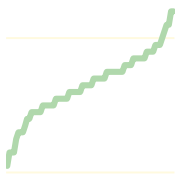
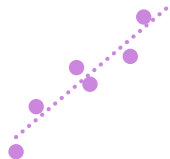
Население мира в возрасте от 60 до 100 лет составило 737 миллионов человек.

05

Корреляционное

Количество бактерий на упавшей на пол конфете возрастает при увеличении времени нахождения этой конфеты на полу.

Выбор типа диаграммы

	ПОКОМПОНЕНТНОЕ	ПОЗИЦИОННОЕ	ВРЕМЕННОЕ	ЧАСТОТНОЕ	КОРРЕЛЯЦИОННОЕ
КРУГОВАЯ					
ЛИНЕЙЧАТАЯ					
ГИСТОГРАММА					
ГРАФИК					
ТОЧЕЧНАЯ					

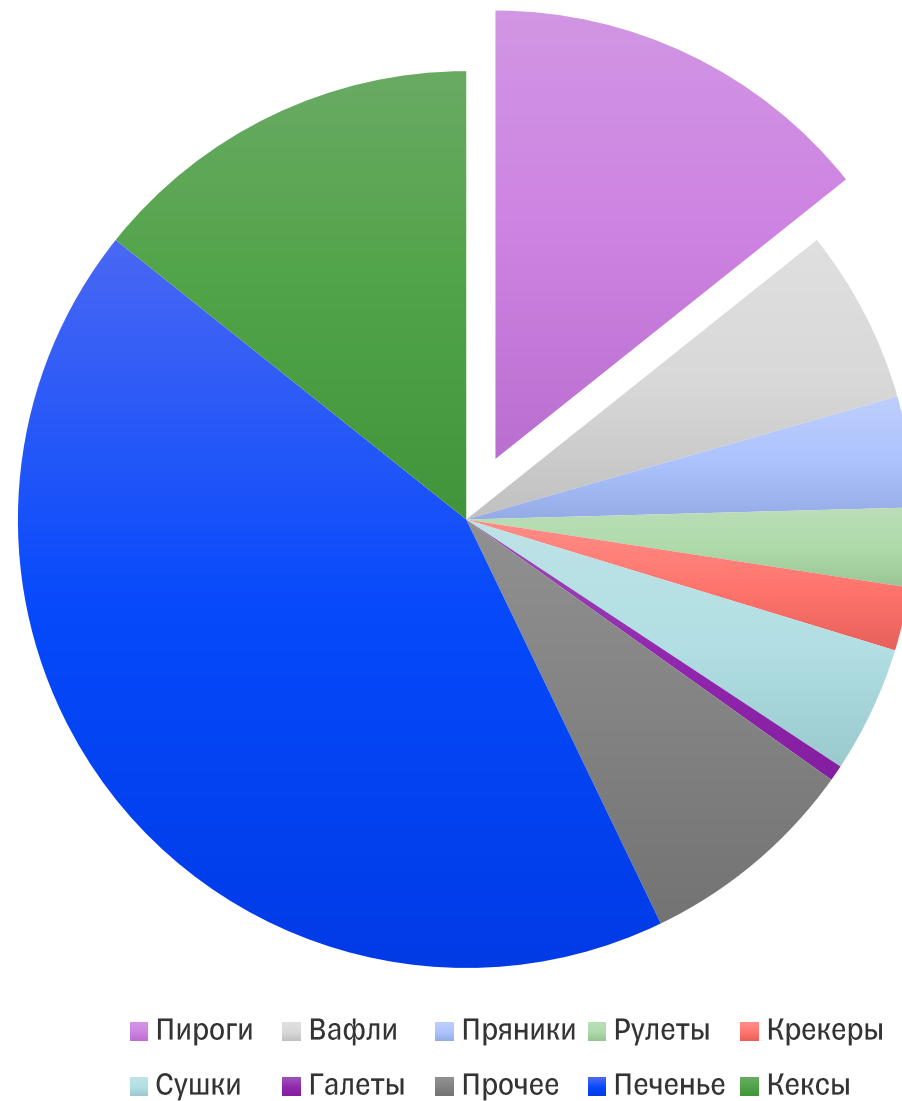


ТИНЬКОФФ

Тривиальные визуализации

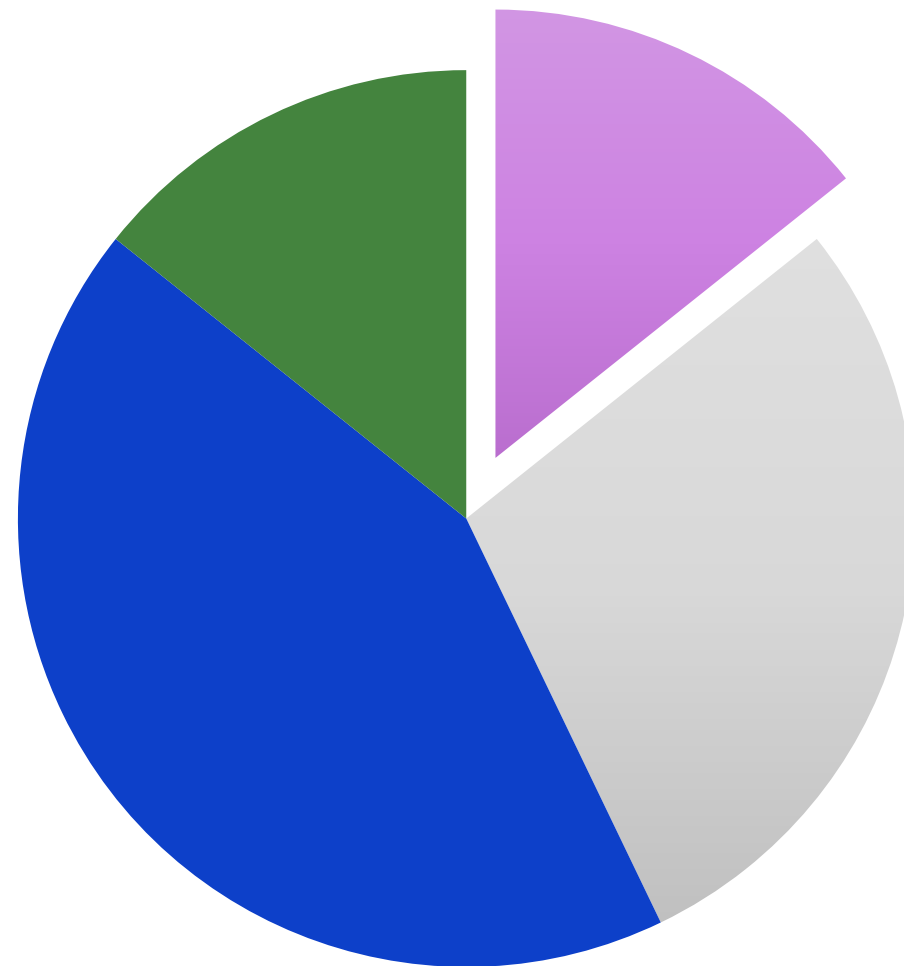
Круговая диаграмма - Pie chart

- Неизвестно, что отражает диаграмма
- Сложно сравнивать мелкие доли
- Нет числовой конкретики



Круговая диаграмма - Pie chart

- Неизвестно, что отражает диаграмма
- Нет числовой конкретики



■ Пироги ■ Прочее ■ Печенье ■ Кексы

Круговая диаграмма - Pie chart



Хорошо подходит для
покомпонентного
сравнения



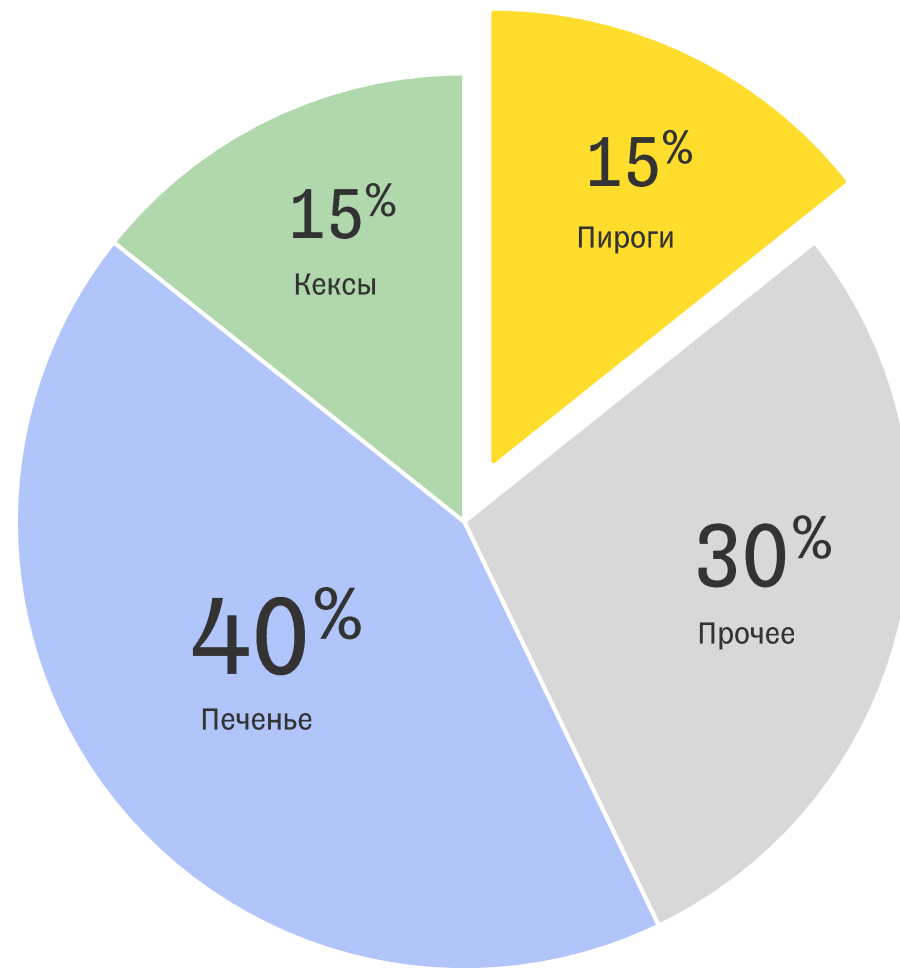
Используется при
отображении одного
ряда данных



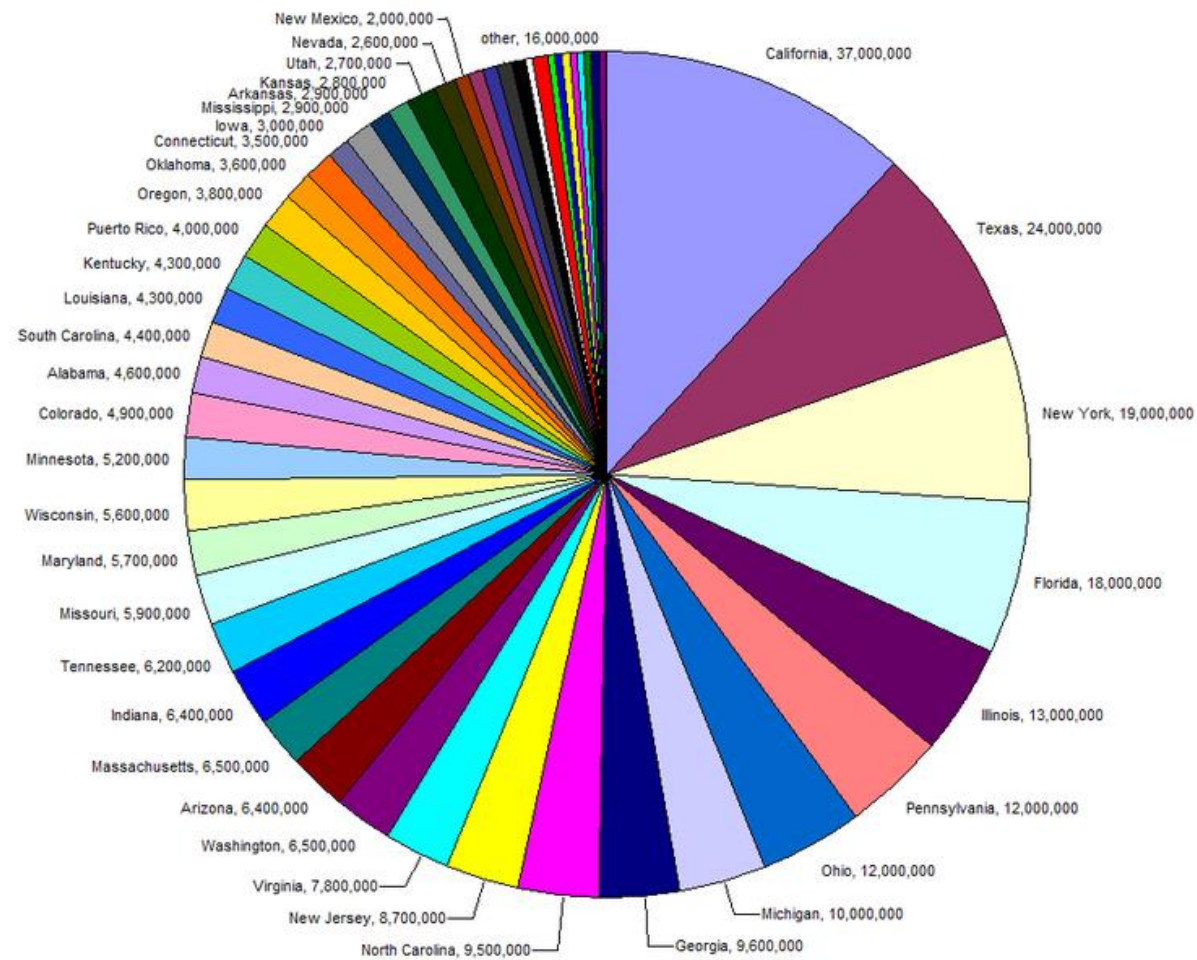
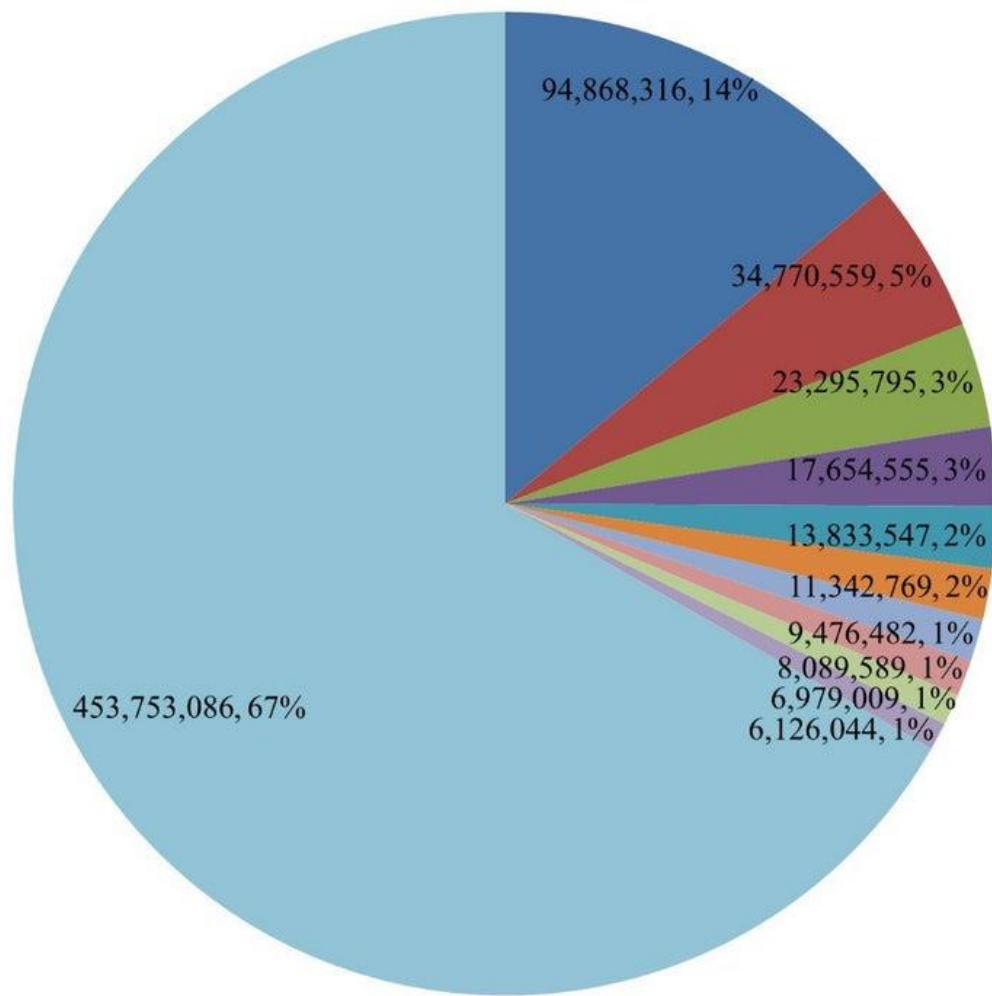
Показывает составные
части одного целого



Не включайте в
диаграмму больше 4-5
категорий

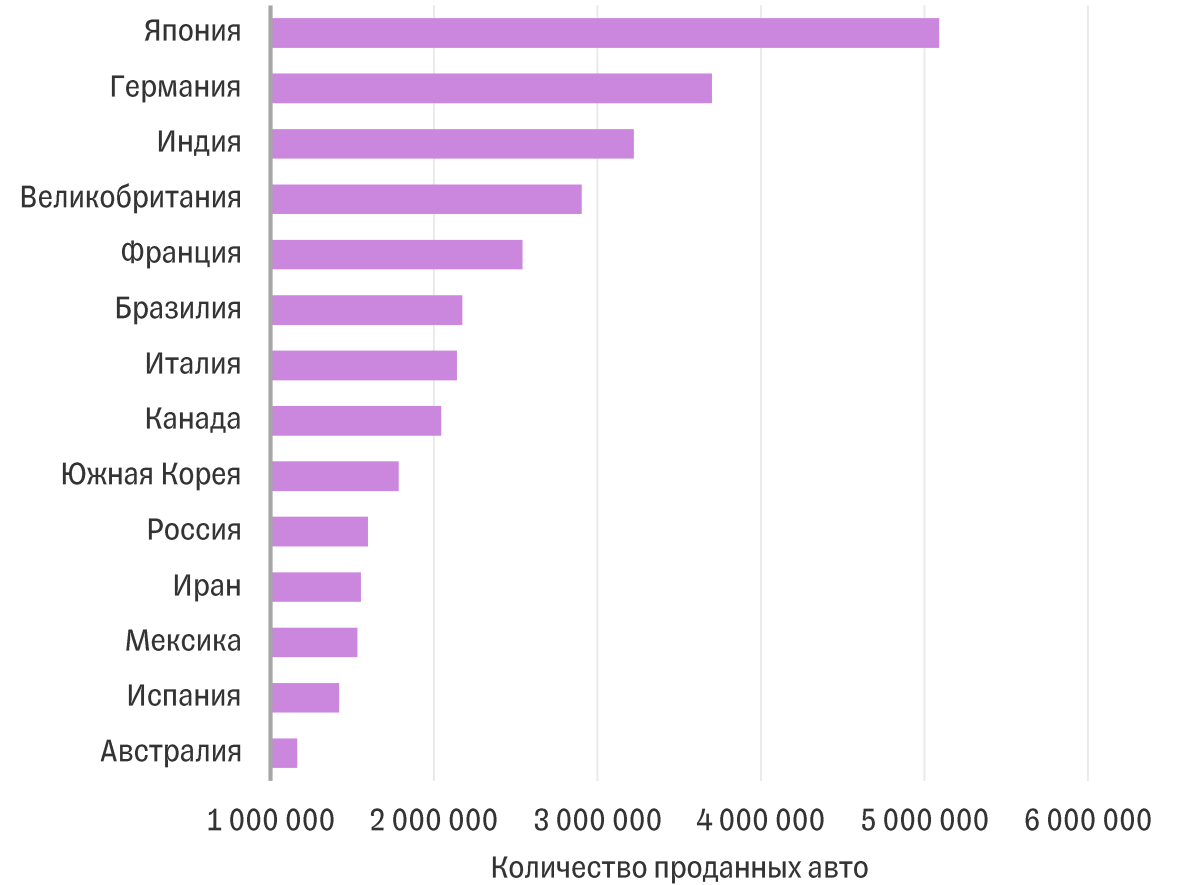


Продажа мучных кондитерских изделий в России



Линейная диаграмма - Bar chart

- Смещенная ось

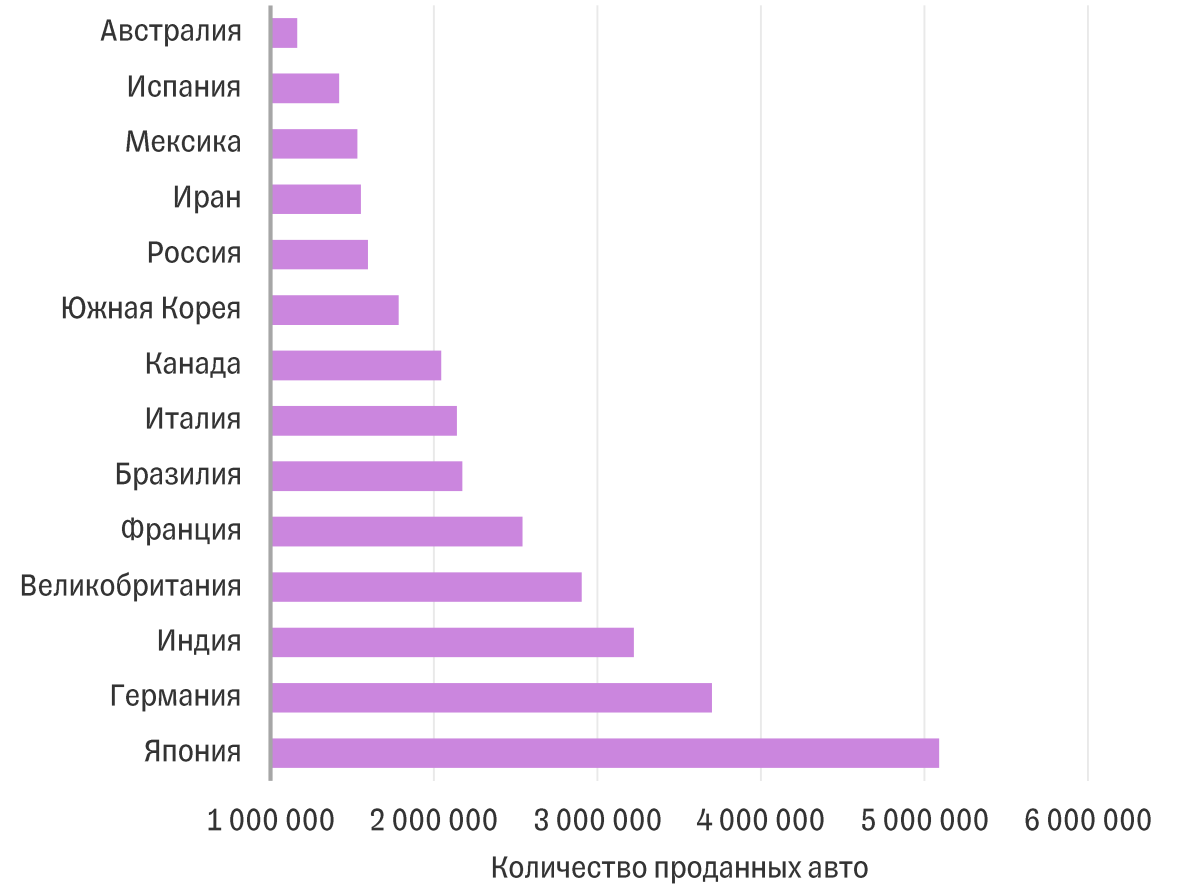


Рейтинг стран мира по продажам авто

*Не включая Китай (30,20%) и США (19,20%)

Линейная диаграмма - Bar chart

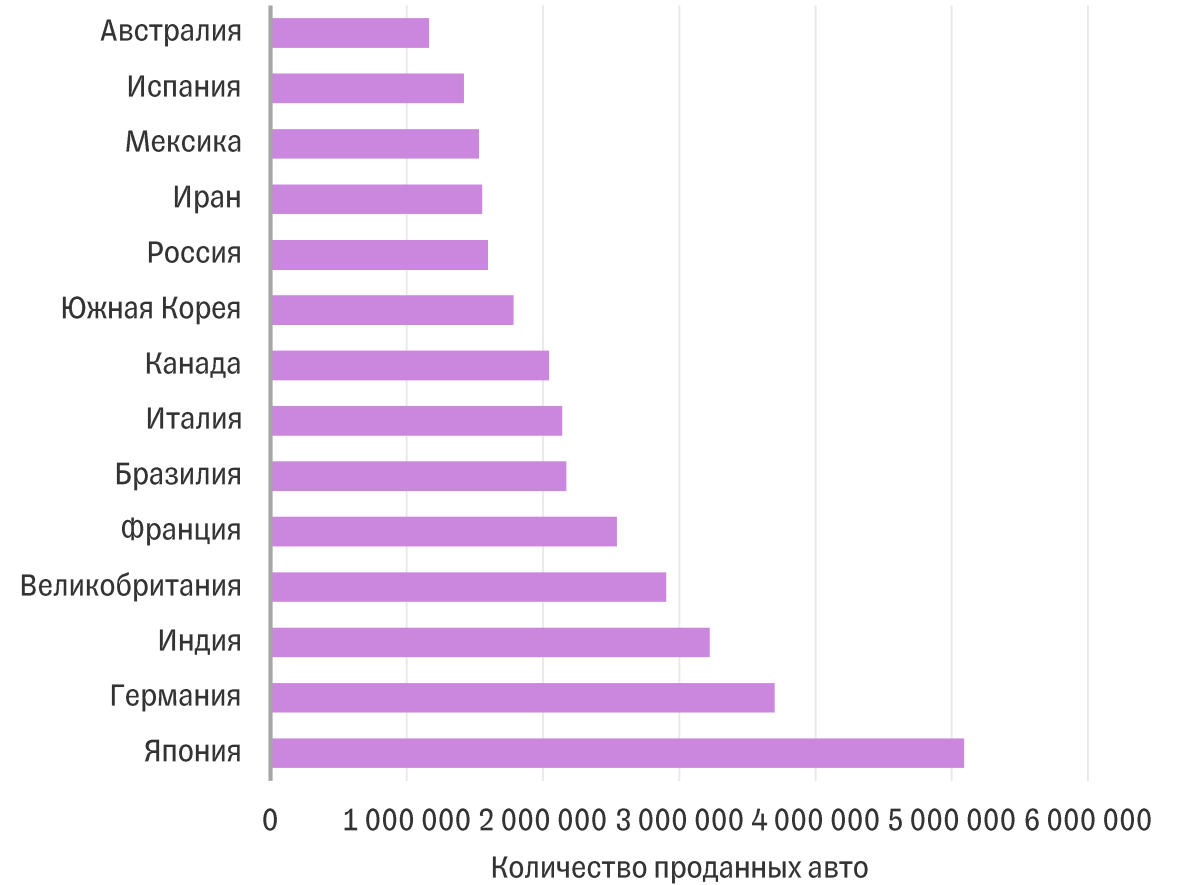
- Смещенная ось



Рейтинг стран мира по продажам авто

*Не включая Китай (30,20%) и США (19,20%)

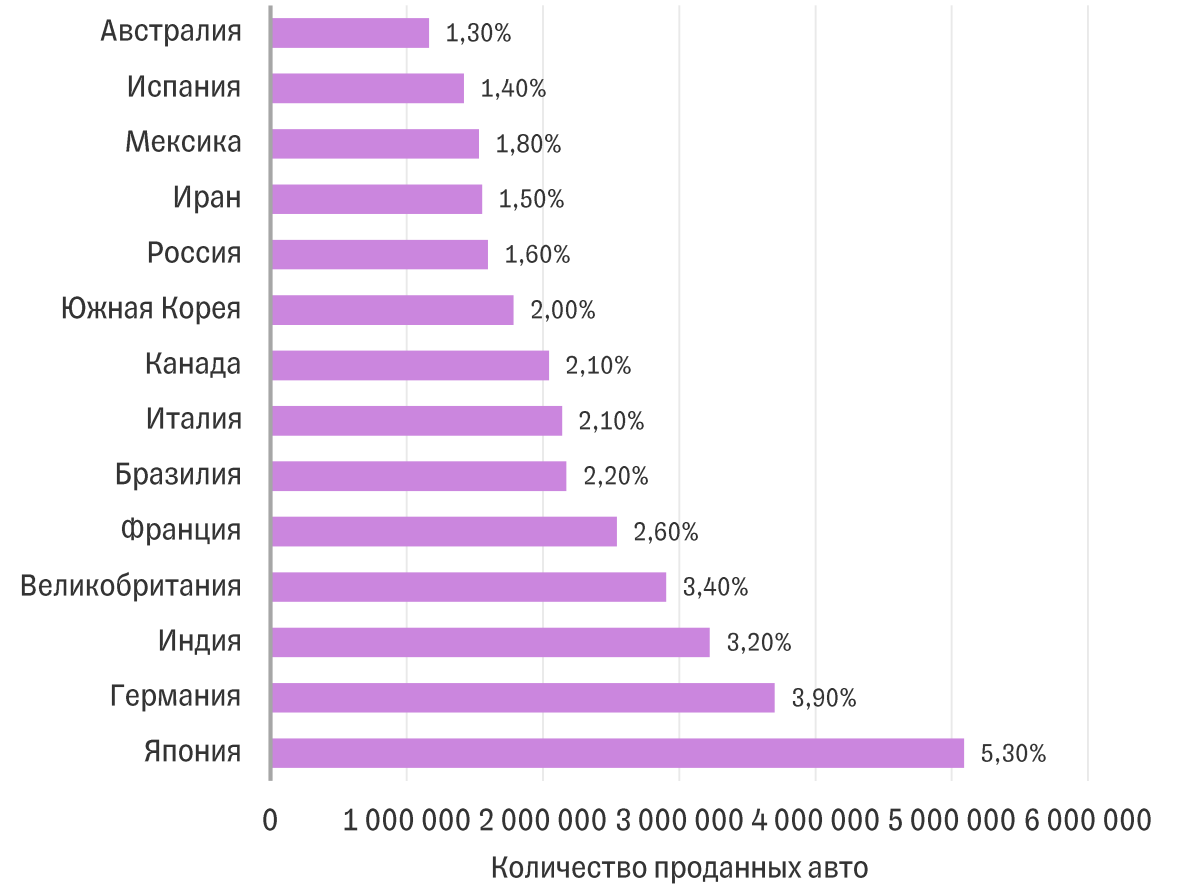
Линейная диаграмма - Bar chart



Рейтинг стран мира по продажам авто

*Не включая Китай (30,20%) и США (19,20%)

Линейная диаграмма - Bar chart



Рейтинг стран мира по продажам авто

*Не включая Китай (30,20%) и США (19,20%)

Линейная диаграмма - Bar chart



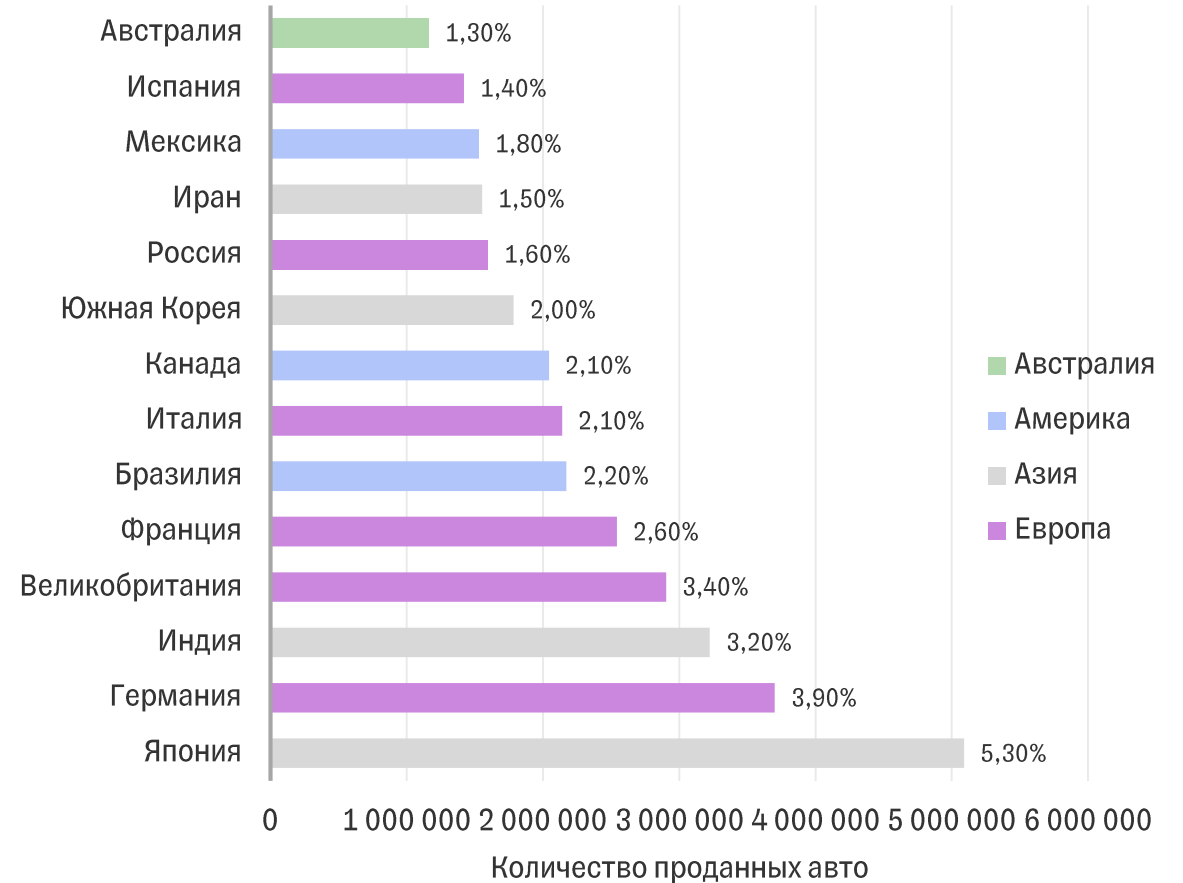
Хорошо подходит для
позиционного
сравнения



Удобно показывает
большое количество
данных



Не изменяйте масштаб
осей, если не хотите
обмануть человека

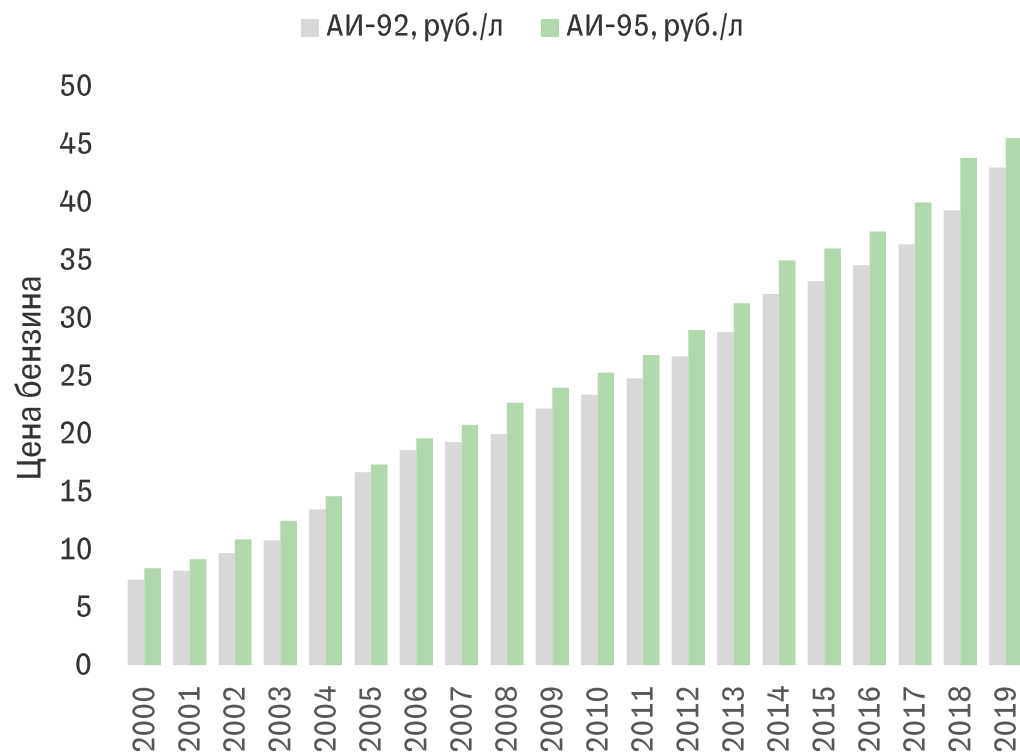


Рейтинг стран мира по продажам авто

*Не включая Китай (30,20%) и США (19,20%)

Гистограмма - Histogram

Таблица роста цен на бензин за последние 20 лет



Хорошо подходит для
временного и
частотного сравнения



Можно смотреть
динамику нескольких
показателей

Различия гистограмм и линейных диаграмм



Сравниваем категории



Сравниваем динамику во
времени



Сравниваем частоту появления
значения в интервале

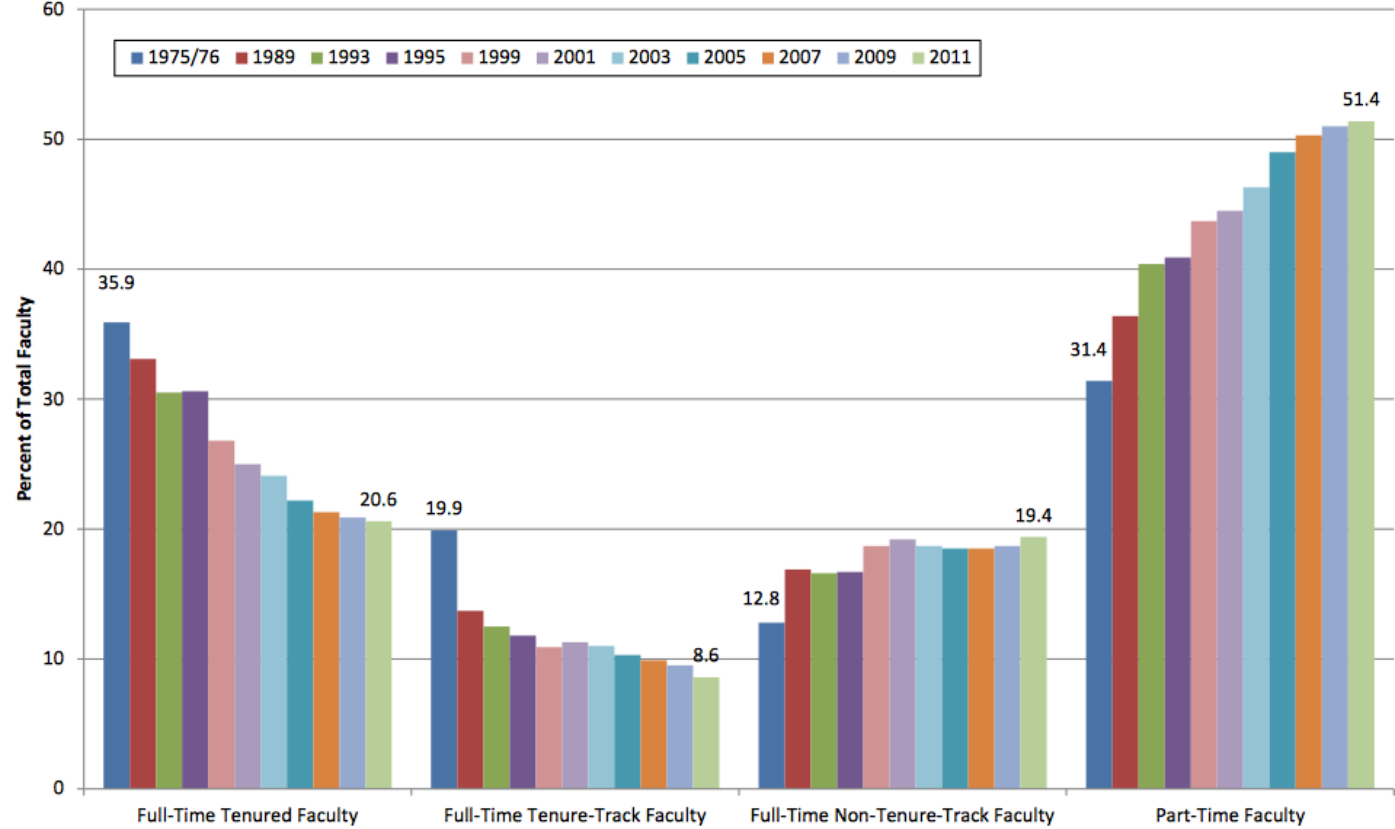


Одна метрика

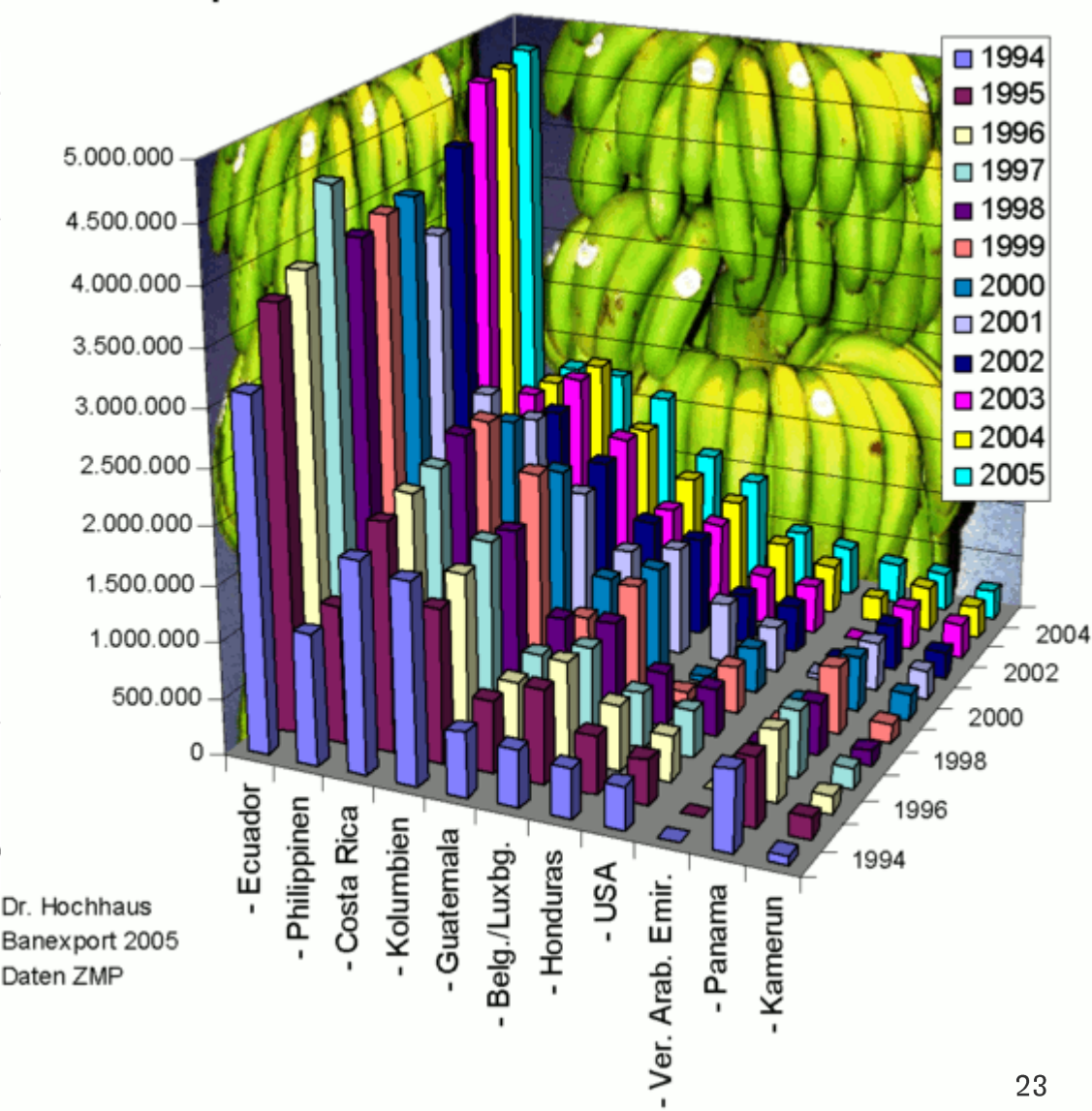


Несколько метрик изображенных
рядом

Trends in Faculty Employment Status, 1975 and 1976 to 2011



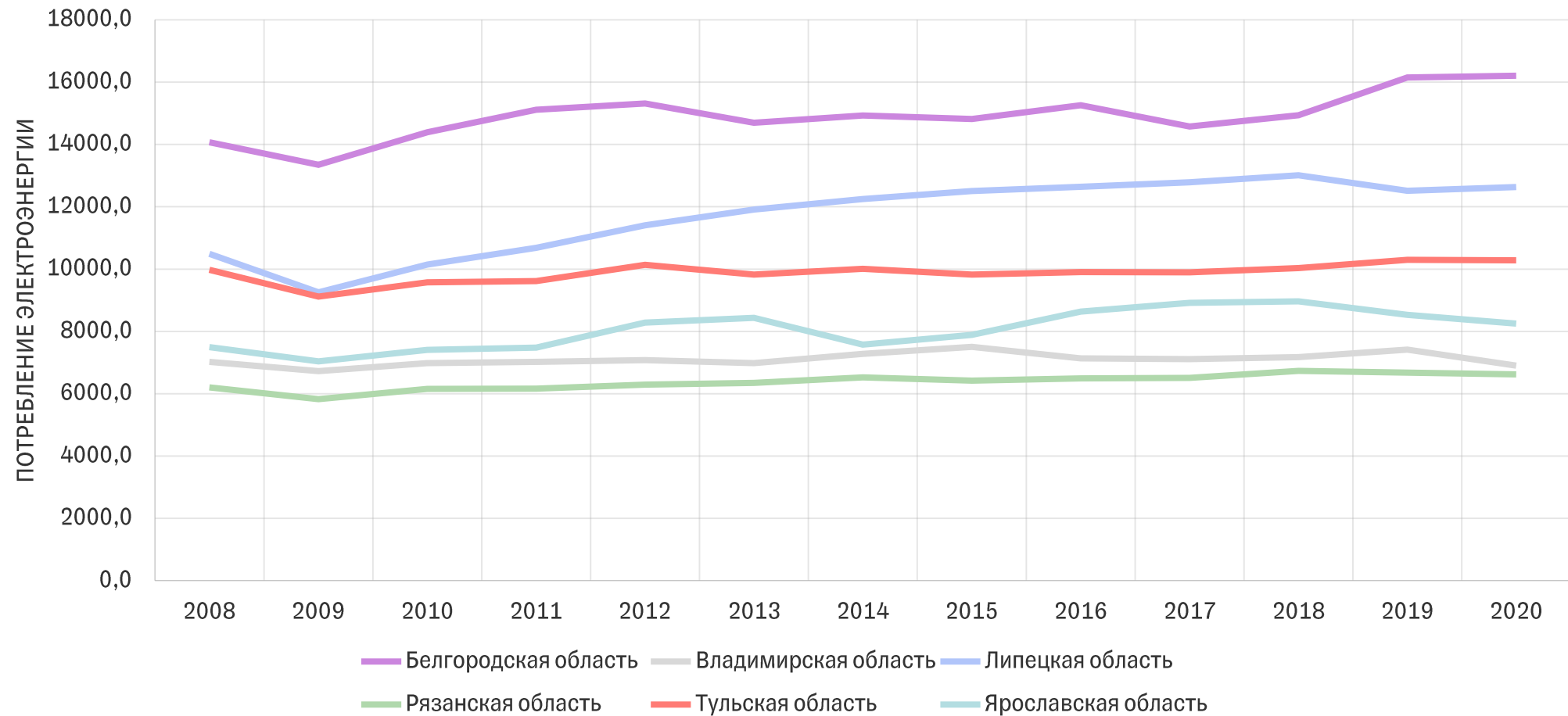
Export von Bananen in Tonnen von 1994-2005

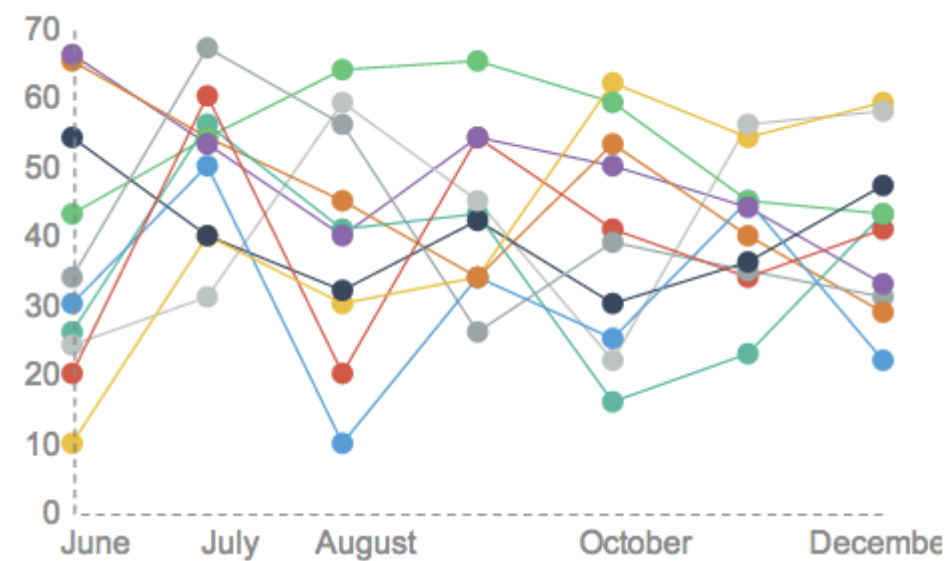
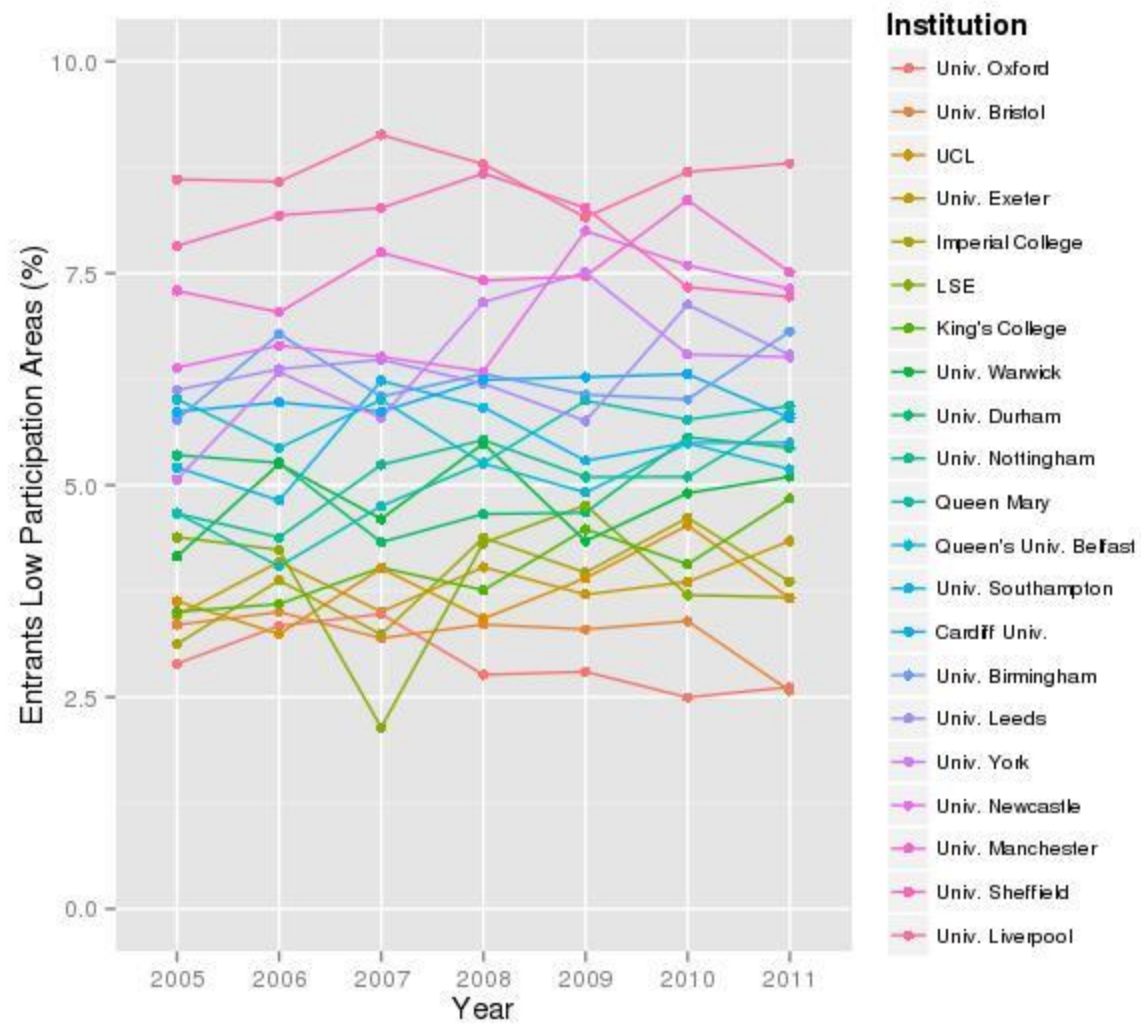


Dr. Hochhaus
Banexport 2005
Daten ZMP

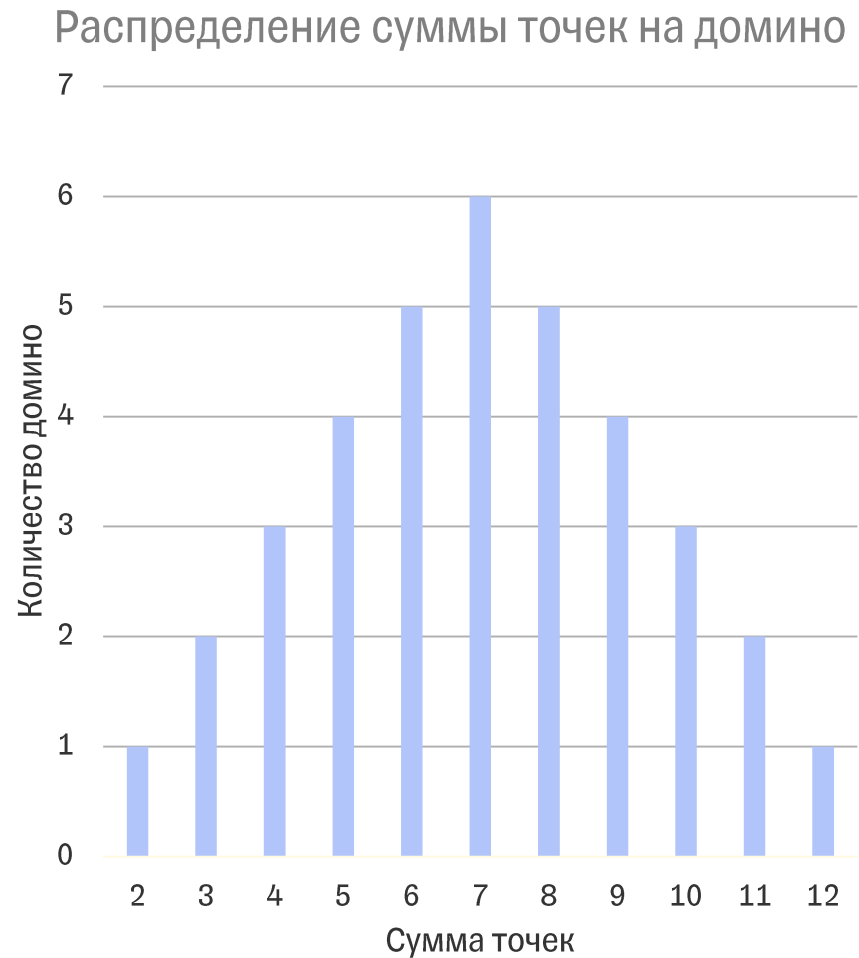
Линейный график - Line plot

Потребление электроэнергии по субъектам Российской Федерации (млн.кВт.час)





И снова гистограмма И снова график



Точечная диаграмма - Scatter plot



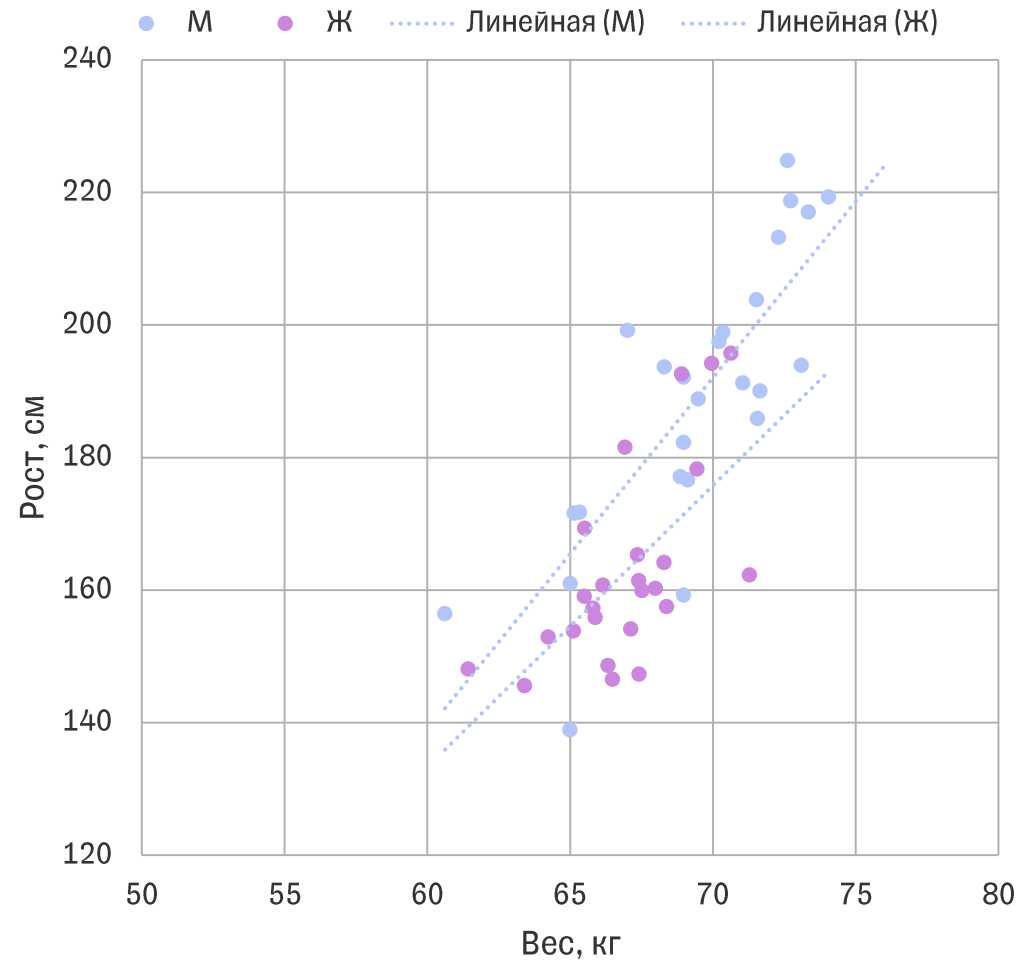
Хорошо подходит для
корреляционного
сравнения



Можно показывать
большое количество
измерений



Концентрация точек
вдоль линии тренда
указывает на
зависимость значений
метрик

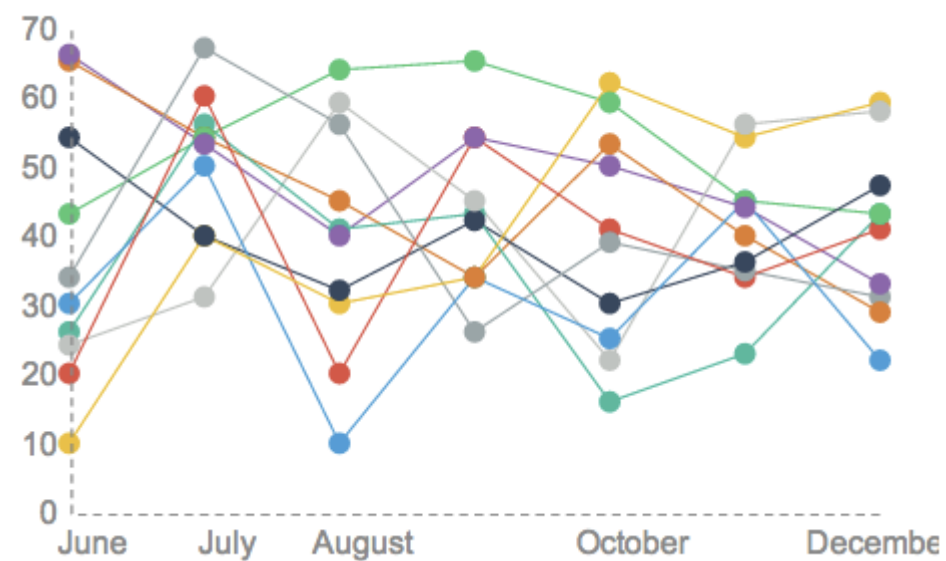
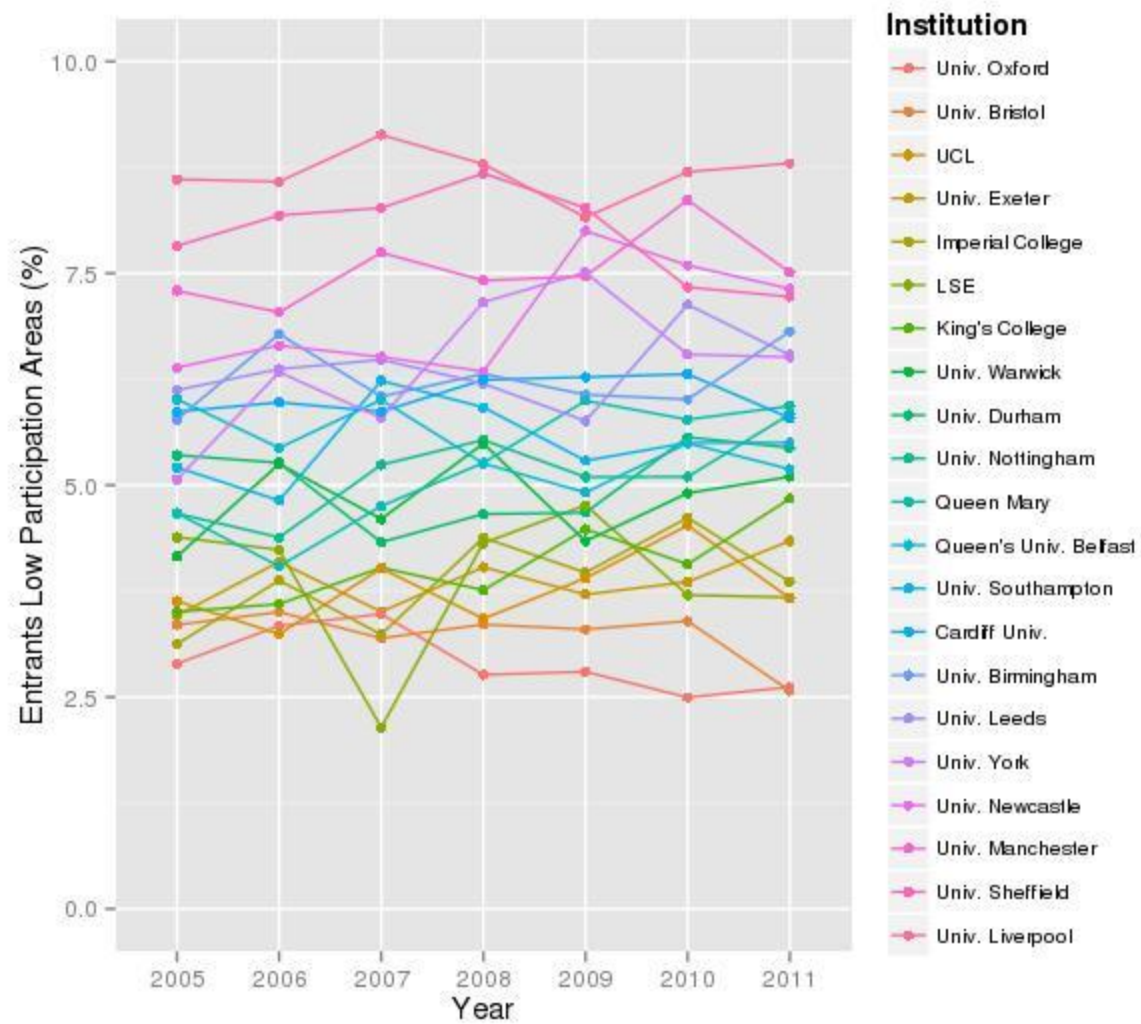


Корреляция веса и роста людей

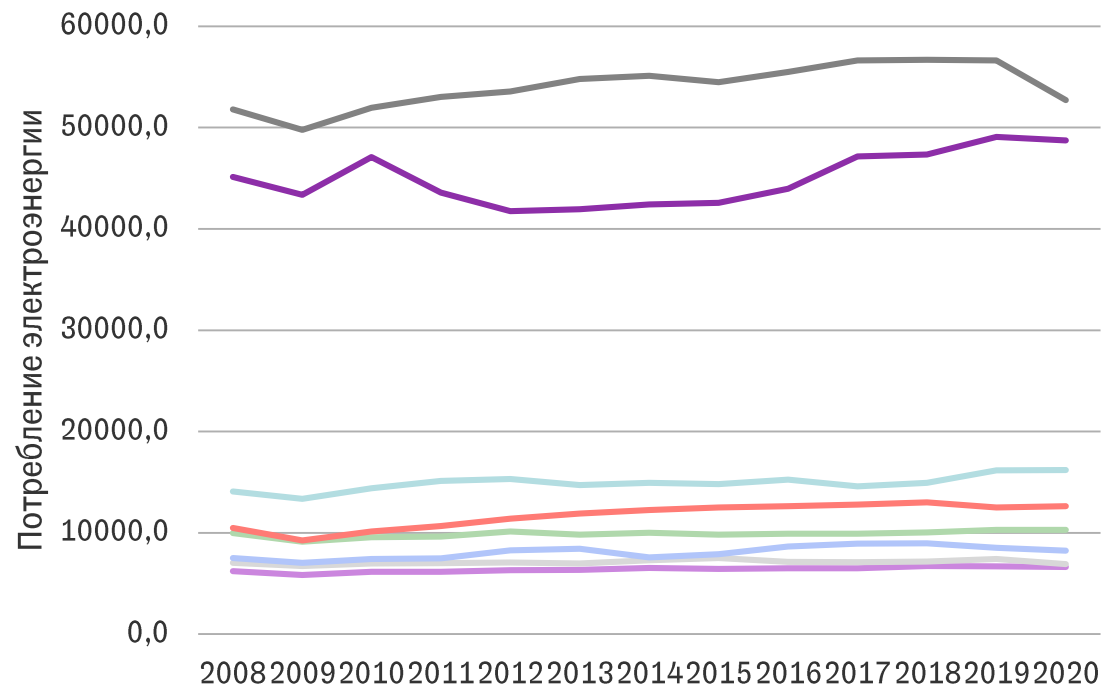


ТИНЬКОФФ

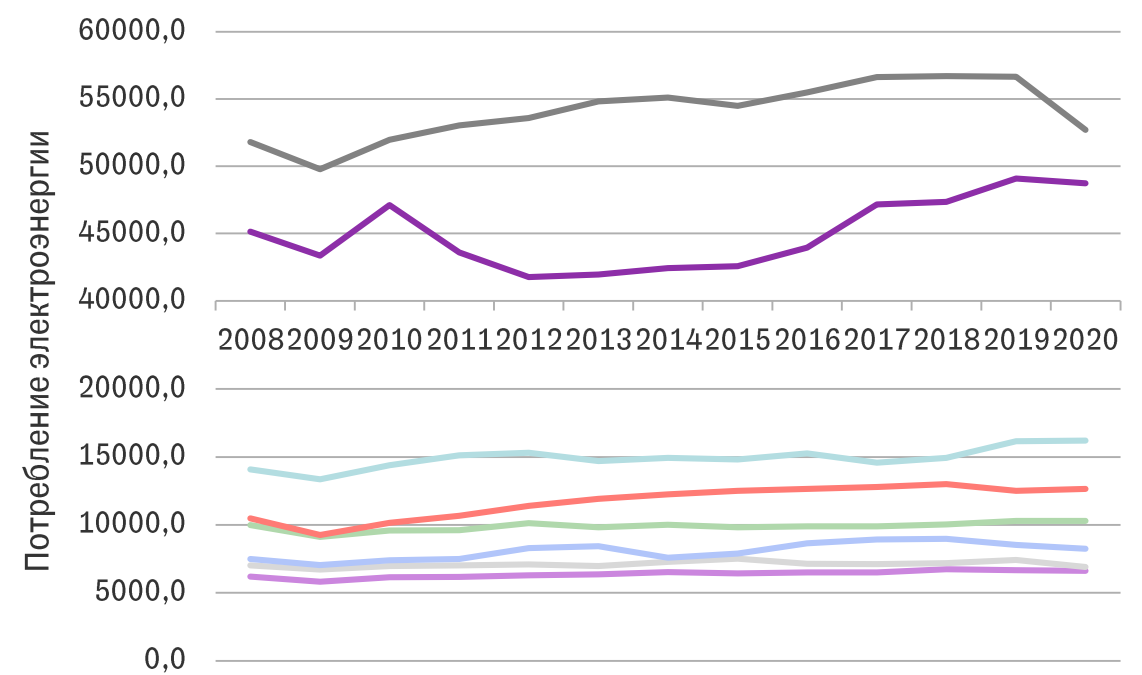
Визуализации с изюминкой



Потребление электроэнергии по субъектам Российской Федерации (млн.кВт.час)



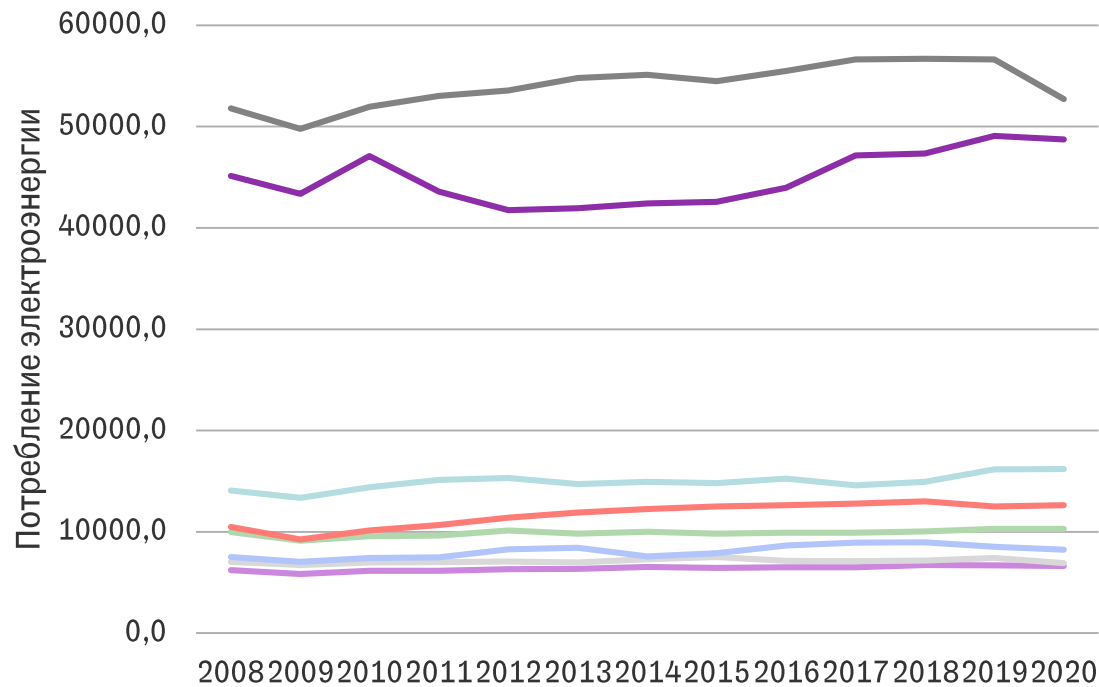
Рязанская область
 Ярославская область
 Липецкая область
 Московская область
 Владимирская область
 Тульская область
 Белгородская область
 г.Москва



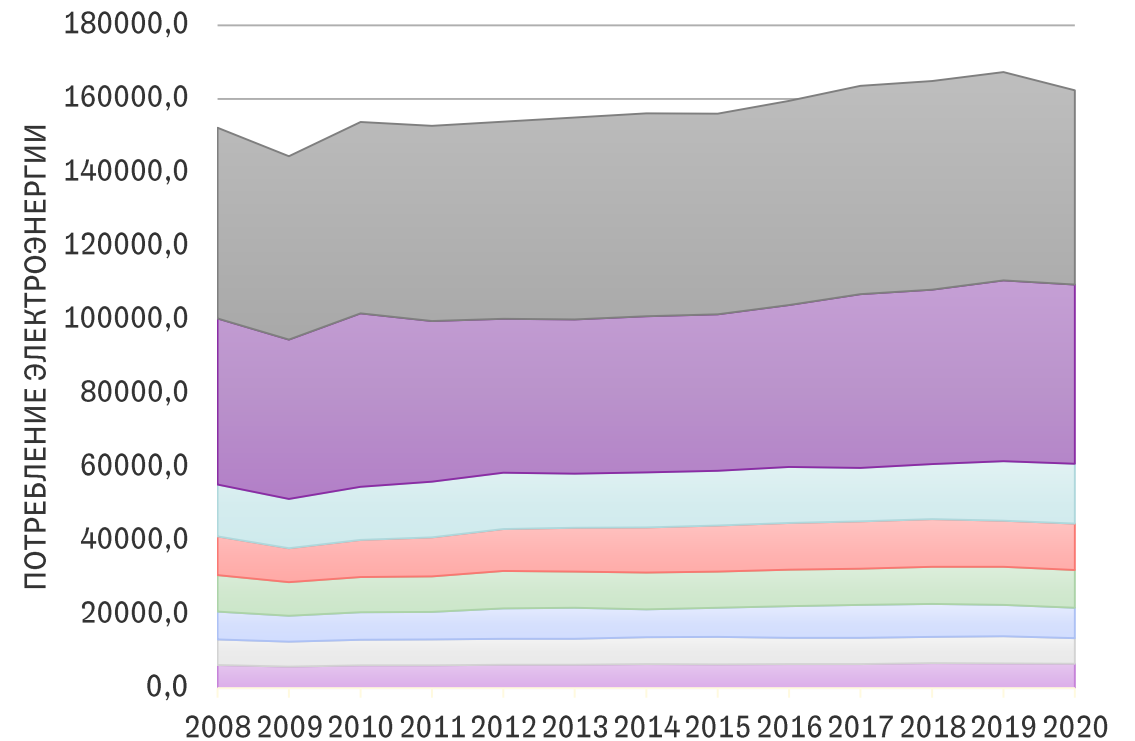
Рязанская область
 Ярославская область
 Липецкая область
 Московская область
 Владимирская область
 Тульская область
 Белгородская область
 г.Москва

Диаграмма с областями - Area chart

Потребление электроэнергии по субъектам Российской Федерации (млн.кВт.час)



Рязанская область Владимирская область
Ярославская область Тульская область
Липецкая область Белгородская область
Московская область г.Москва

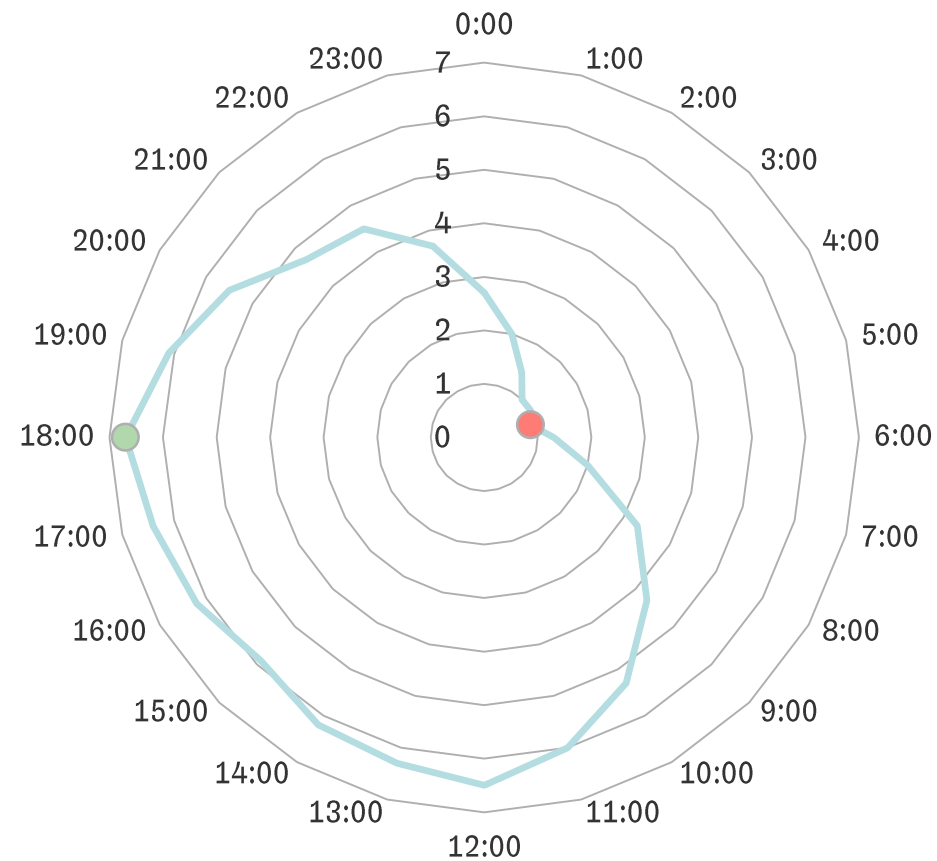


Рязанская область Владимирская область Ярославская область
Тульская область Липецкая область Белгородская область
Московская область г.Москва

Лепестковая диаграмма - Radar chart



Хорошо подходит для
циклического
временного сравнения



Активность аудитории Telegram по времени суток

Тепловая карта - Heat map

Продажи солнцезащитного крема

	Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
2011	5	2	20	36	60	101	99	60	21	12	5	5
2012	11	2	22	43	70	112	85	63	20	2	4	3
2013	7	10	24	30	76	99	104	47	26	10	4	10
2014	1	13	20	31	52	85	102	57	23	8	7	15
2015	4	9	18	29	65	100	98	49	18	5	11	12
2016	6	6	18	38	42	117	100	64	22	6	10	9
2017	7	6	22	45	77	112	122	69	13	5	11	4
2018	4	5	19	35	74	101	119	64	27	12	1	8
2019	11	2	15	32	73	93	99	54	17	11	1	1



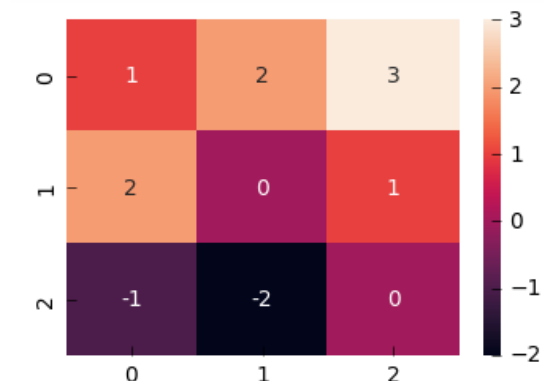
Хорошо подходит для
позиционного
сравнения



Применяется при
проверке корреляции
между метриками



Можно смотреть
динамику показателя в
нескольких циклах



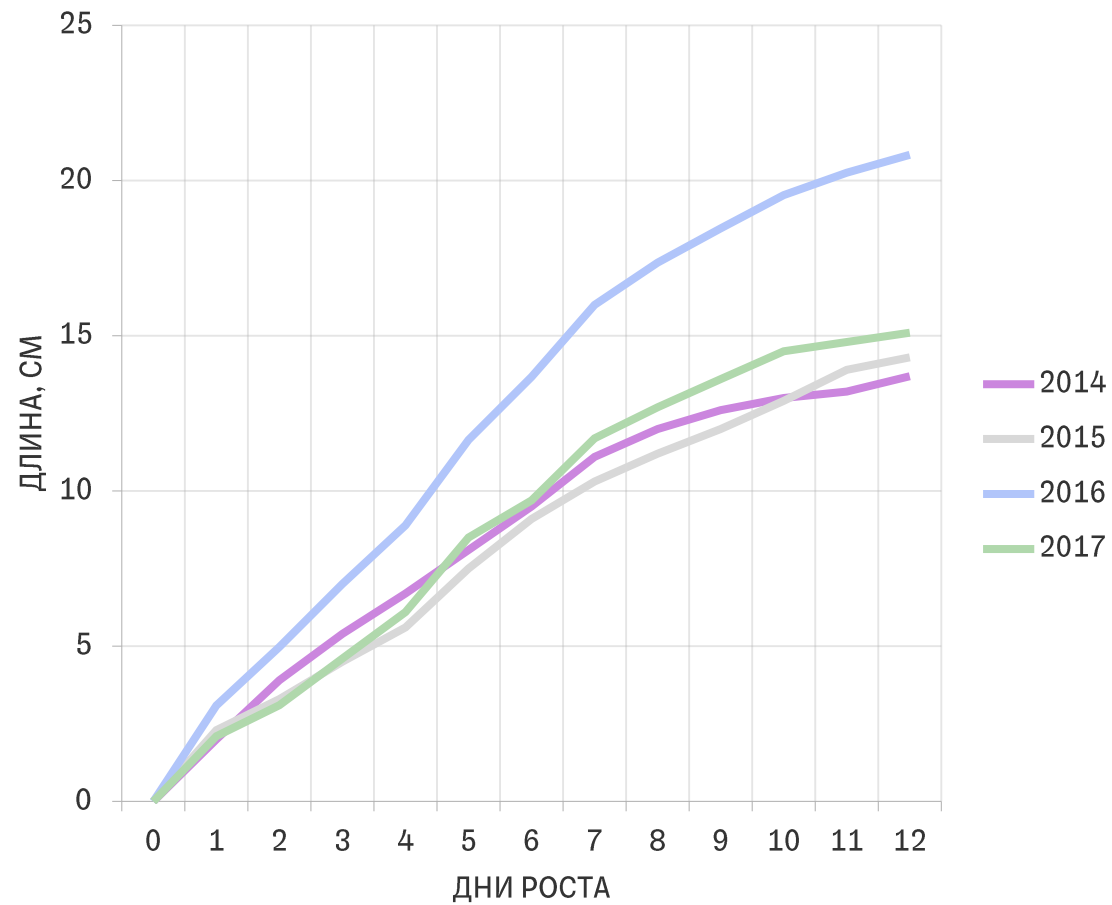


ТИНЬКОФФ

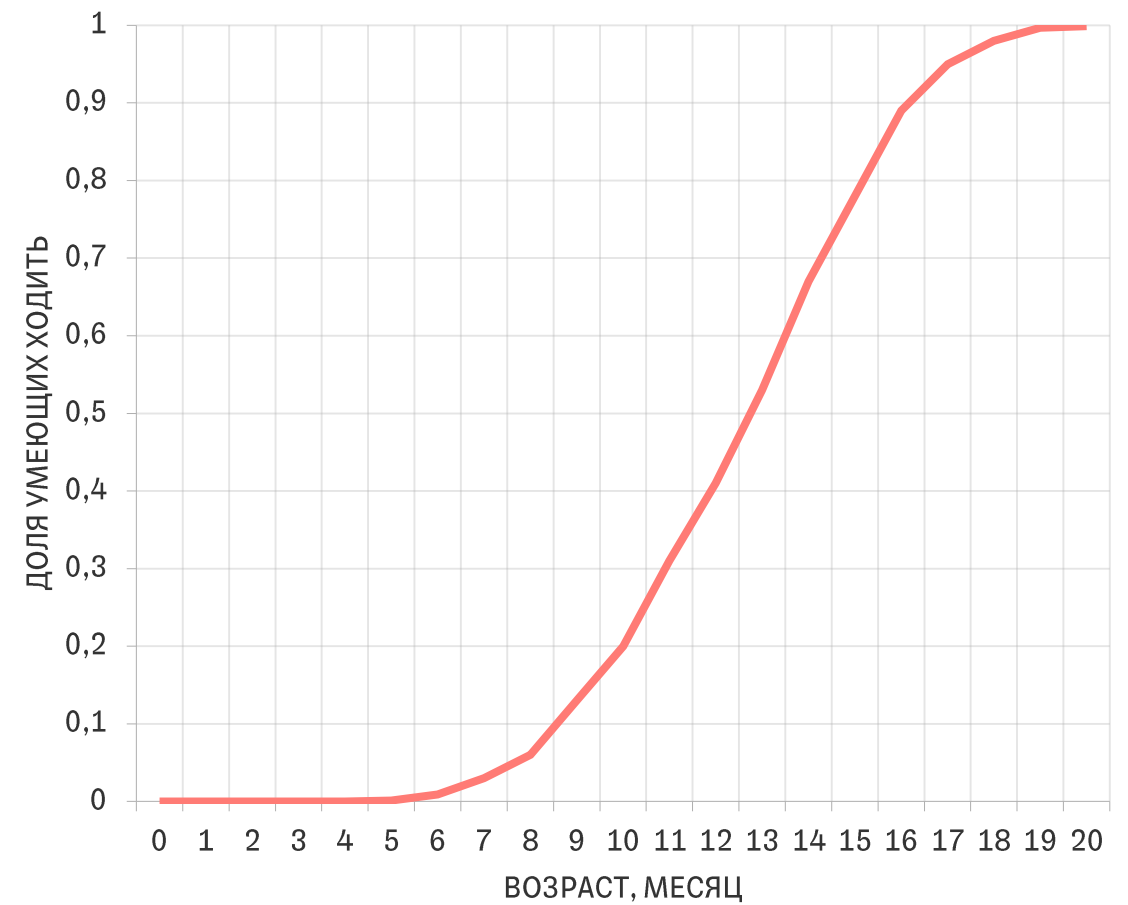
Кексикі с изюмом среди
визуализаций

Винтажный анализ - Vintage analysis

Винтаж роста огурца



Винтаж развития ходьбы у детей



Ящик с усами (диаграмма размаха) - Box plot



Границы ящика
показывают 25 и 75
перцентили



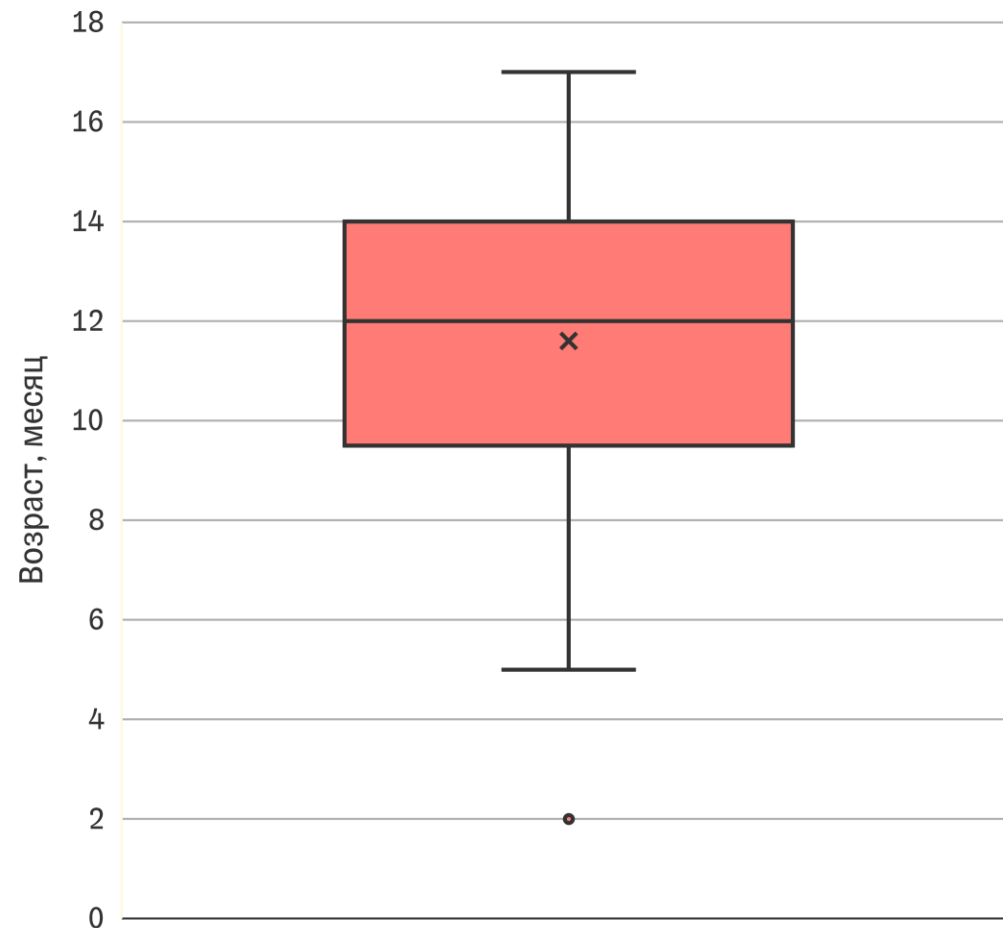
Буквы Т показывают
минимальное и
максимальное
значения



Черта внутри ящика
показывает медиану



Маленький крестик
показывает среднее
арифметическое



Распределение возраста, когда дети начинают
ходить



ТИНЬКОФФ

Google в помощь



ТИНЬКОФФ

ИТОГИ

Виды сравнения метрик

01

Покомпонентное

Круговая диаграмма
Диаграмма с
областями

02

Позиционное

Линейная диаграмма
Гистограмма
Диаграмма с
областями
Тепловая карта

03

Временное

Гистограмма
График
Диаграмма с
областями
Лепестковая
диаграмма
Тепловая карта
Винтаж

04

Частотное

Гистограмма
График
Ящик с усами

05

Корреляционное

Точечный график
Тепловая карта

На что обращаем внимание



Выбрать подходящий тип графика



Выбрать подходящий инструмент



Убрать все лишнее (ненужные данные, выбросы)



Выбрать масштаб



Смотреть на совокупность метрик (разных по сути, разных по типу)



Смотреть на динамику

Не допускаем ошибки



Подписи графика и осей



Выделение важного параметра



Подбор информативной
визуализации



Спокойные тона при выборе
цвета

Мелкие неразличимые доли
диаграмм

«Лапша» из графиков

Смещение осей диаграммы

«Бразильский карнавал» в
палитре цветов



ТИНЬКОФФ









Джин Желязны

«Говори на языке диаграмм:
пособие по визуальным
коммуникациям»



ТИНЬКОФФ

Excel: плюсы и минусы

- | | |
|--|---|
|  Быстро и легко строятся графики |  Тяжело работает с большими таблицами |
|  Широкий простор для форматирования |  Нет автоматизации |
|  Вычисление по формулам без кода |  Нет подключения к базам данных |
|  Дружит с Outlook и PowerPoint |  Нет наглядности: по каким правилам собраны данные |



ТИНЬКОФФ

Python: плюсы и минусы



Автоматизация



Большое количество библиотек
для визуализации



Есть подключение к базе



Удобно обмениваться кодом



Высокая скорость обработки



Графики сложнее настроить



Нет простой интеграции
с продуктами Office



Нельзя дать менеджеру данные
«покрутить»



ТИНЬКОФФ

Tableau: плюсы и минусы



Отчеты интерактивные



Подходит для управленческой отчетности



Есть подключение к базе



Интуитивно понятный интерфейс



Инструмент платный



Если данных много, возможны зависания



ТИНЬКОФФ

Он такой один